基于 CARLA-PSO 组合模型的智能控制器参数学习优化 *

谷学静 [1]，张明儒 [1]，王志良 [2]，郭宇承 [3]

(1. 华北理工大学 电气工程学院，河北 唐山 063009; 2. 北京科技大学 计算机与通信工程学院，北京 100083; 3. 华北理工大学 轻工学院，河北 唐山 063000)

摘　要：对连续动作强化学习自动机（CARLA）进行了改进，应用改进后的 CARLA 结合粒子群优化算法（PSO）优化 PID 参数。以 CARLA 为基础，建立了 CARLA 和 PSO 的组合优化学习模型 CARLA-PSO，该模型包含 CARLA 学习环路和 PSO 学习环路两个部分，通过优化策略选择器进行学习环路的选择，通过与环境进行相互作用，获得最优控制。本文对连铸结晶器液位控制进行了仿真实验，实验结果表明，CARLA-PSO 在进行 PID 参数优化时寻优效率高，全局搜索能力强，能够达到理想的控制效果。具有较好的应用前景。

关键词：连续动作学习强化自动机；粒子群优化算法；智能 PID 控制器；结晶器液位

中图分类号：TP301.6　　doi: 10.3969/j.issn.1001-3695.2017.09.0927

Parameter learning optimization of intelligent controller based on CARLA-PSO Composite Model

Gu Xuejing[1], Zhang Mingru[1], Wang Zhiliang[2], Guo Yucheng[3]

(1. *School of Electrical Engineering*, *North China University of Science & Technology*, *Tangshan Hebei* 063017, *China*; 2. *School of Computer & Conmunication Engineering*, *University of Science & Technology Beijing*, *Beijing* 100083, *China*; 3. *College of Light Industry*, *North China University of Science & Technology*, *Tangshan Hebei* 063000, *China*)

ABSTRACT This paper presented a hybrid approach involving continuous action reinforcement learning automata (CARLA) and particle swarm optimization (PSO) to design a optimal and intelligent proportiona-lintegral-derivative (PID) controller of an mould level control system. The proposed method is CARLA which is able to explore and learn to improve control performance without the knowledge of the analytical system model. CARLA-PSO is composed of two sections which are the CARLA learning loop, and the PSO learning loop, and select learning loop according to the Optimal policy selector. CARLA-PSO is a method that combines the features of PSO and CARLA in order to improve the optimize operation through interaction with the environment. The experimental results show that CARLA-PSO was indeed more efficient in improving the step response of an mould level control system.

Key Words: Continuous action reinforcement learning automata; particle swarm optimization; intelligent PID controller; mould level

0　引言

结晶器是连铸机上的重要设备，其液位的控制精度对提高生产率、生产特殊钢材、降低连铸的漏钢事故率有重要意义[1]。由于结晶器液位控制系统具有非线性、强耦合等特点，且扰动因素较多[2]，传统的 PID 控制只在特定工艺条件下才能实现较好的液位闭环控制。

智能控制理论是对人或动物局部智能的模仿，已成为解决复杂与难控系统的主要途径。针对传统 PID 控制方法在结晶器液位控制中的不足，童朝南[3]等将一种基于遗传算法的有约束广义预测控制方法应用于结晶器液位控制，控制效果优于传统 PID 控制。为了实现结晶器液位 PID 控制器参数的自整定，曹光明[4]等提出了模糊自适应 PID 控制策略；房启超[5]等提出了基于变惯性性权重和多种群并行优化策略的改进 PSO 算法。

PSO 是一种基于种群的随机优化技术，被广泛用于解决多目标、多参数的学习优化问题[6-8]。强化学习（RL）是一种在线

基金项目：国家自然科学基金资助项目（61170117）；国家重点研发计划资助项目（2016YFB1001404）；河北省自然科学基金高端钢铁冶金联合研究基金专项项目（F2017209120）

作者简介：谷学静（1972-），女，教授，博士，主要研究方向为冶金自动化、人机交互（854169781@qq.com）；张明儒（1991-），男，硕士，主要研究方向为冶金自动化；王志良（1956-），男，教授，博士，主要研究方向为人工心理数字化技术与研究、解说型服务机器人、冶金自动化等；郭宇承（1968-），男，副教授，硕士，主要研究方向为工业设计、产品交互.

的，介于监督学习和非监督学习之间的机器学习方法[9]，已被用于导弹制导控制与飞机的智能逃避 [10]、机器人在未知环境下的自动避障[11]、学习型双连杆机械臂的控制[12] 等智能控制领域。学习自动机（LA）是强化学习的主要工具[13]，CARLA 是其中的一种，其主要特点是可以输出连续的动作，具有很强的全局搜索能力。MN Howell[14]等人将 CARLA 方法应用于发动机怠速系统 PID 控制器的参数整定，在系统模型信息不足的情况下,有效的解决了车辆尾气排放超标和汽车失速的问题；M Kashki[15]等人将 CARLA 方法应用于功率同步发电机自动电压调节器 PID 参数自整定，仿真结果证明，与 PSO、遗传算法相比，该方法具有更高的控制精度。

本文以 CARLA 为基础设计了 CARLA-PSO 组合优化模型，该组合优化模型优化过程分为两个阶段，每个阶段分别对应着 CARLA 学习环路、PSO 学习环路。前期阶段采用 CARLA 进行“粗调”优化，使环境适应度评价值迅速收敛到稳定状态；后期阶段采用 PSO 进行“微调”优化，进一步提高环境适应度评价值的精度。该模型在连铸结晶器液位控制器的参数在线学习优化中表现良好。

1　CARLA-PSO 组合模型的设计

本文建立了一种 CARLA-PSO 组合模型，用于智能控制器参数在线学习优化。如图 2 所示，通过 “策略选择器”进行学习环路选择。启动“策略选择器”，利用 CARLA 学习环路进行全局搜索，当环境适应度值将为学习环路启动时的 10%时，“策略选择开关”拨至 PSO 学习环路继续搜索最优参数。具体步骤如下：

a)初始化参数，例如：迭代次数、种群个数、粒子范围等。

b)通过 CARLA 学习环路确定 PSO 学习环路粒子初始位置。

c)计算每一代的环境适应度值。

d)计算每一代每个粒子群的局部最优位置和全局最优位置。

e)更新粒子群速度、位置、局部最优位置、全局最优位置。

f)重复步骤 c)~e)，直到达到最大代数或连续 m 代环境适应度值不在发生变化。

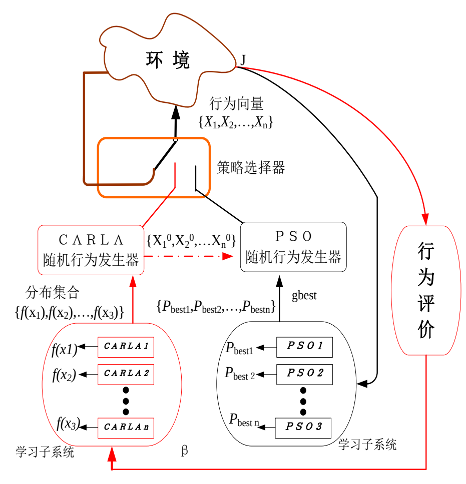


图 1　CARLA-PSO 组合模型学习环路

1.1　CARLA 学习环路设计

CARLA 通过试错、随机行为选择与未知或随机环境交互获得策略的改进，应用于连续变化参数的在线寻优。与常用的强化学习不同的是，CARLA 采用连续行为空间取代离散行为空间，通过概率密度函数获得收敛信息，因此 CARLA 更适合于具有连续特征的工程应用当中。

每个学习变量对应一个 CARLA，每个参数的学习过程独立运行，将它们联系起来的唯一因素是动力学环境，并且它们共享代价函数。在每个自动机中，每个行为 x_i 都对应着一个概率密度函数 $f(x_i)$ 作为行为选择的依据。借助强化信号 β 产生表现更好的行为集合，通过学习子系统增大该行为集合被重新选择的概率。CARLA 学习环路的设计主要有以下几步：

a）初始化密度函数。采用均匀分布的形式初始化概率密度函数，即根据每个参数的范围，其最大值与最小值之差的倒数作为概率密度函数，具体公式如下：

$$f^0(x)=\begin{cases}\dfrac{1}{x_{\min}-x_{\max}} & x\in[x_{\min,}x_{\max}]\\ 0 & 其他\end{cases} \tag{1}$$

其中：$x=\{x_1,x_2\cdots x_n\}$ 为决策行为，$f=\{f_1,f_2\cdots f_n\}$ 对应每一个行为的密度函数。

b)行为选择。

$$\int_0^{x_i(k)} f_i(x_i,k)dx_i=z_i(k) \tag{2}$$

其中：随机数 z 取值范围（0,1），将行为集合输入到环境中，通过预先定义好的代价函数得到代价值 $J(k)$

c)行为表现评估计算。得到代价值 $J(k)$ 后，根据下面的公式评估行为表现评估：

$$\beta(k)=\min\left\{\max\left\{0,\ \frac{J_{med}-J(k)}{J_{med}-J_{\min}}\right\},\ 1\right\} \tag{3}$$

其中：通过当前代价值 $J(k)$ 与历史代价值集合中最小值 $J_{\min}$、中位值 J_{med} 的比较，得当前行为的好坏权重。其中行为表现强化信号 $\beta\in[0,1]$，越接近 1，说明行为表现越好。越接近 0，说明行为表现越差。

d)更新概率密度函数。

$$f^{(k+1)}(x)=\begin{cases}\alpha^k(f^{(k)}(x)+\beta(k)H(x,r)) & x\in[x_{\min},x_{\max}]\\ 0 & 其他\end{cases} \tag{4}$$

其中：$H(x,r)$ 为高斯近邻函数，其值可以改变行为的可能性。

$$H(x,r)=\lambda\exp\left(-\frac{(x-r)^2}{2\sigma^2}\right)$$

$$\alpha^{(k)}=\frac{1}{\int_{x_i^{\min}}^{x_i^{\max}}[f_i(x_i,k)+\beta(k)H(x_i,r)]dx_i}$$

$$\lambda=\frac{g_h}{(x_{\max}-x_{\min})}$$

$$\sigma=g_w(x_{\max}-x_{\min})$$

其中：g_h，g_w 分别为高斯分布函数的高度和宽度，其值可以决

定学习的速度；r 为本次学习得到的动作值。α 称为归一化因子，可使密度函数 f 保持在 $[0,1]$ 内。

1.2　一种改进的 CARLA

CARLA 的动作集是一个有限的区间，并采用一种非参数化的概率模型。其初始动作概率为均匀分布，在学习过程中一个对称的高斯近邻函数，将表现较好的动作的奖赏“传播”给相邻动作。由于使用非参数化的概率模型。该方法对概率分布的计算、存储和更新相当复杂，同时在每次学习中都必须进行积分函数的更新，积分方程的求解，计算代价很高，算法的实时性差。

为克服现有算法的不足，本文对 CARLA 做了一些改进，采用式(5)更新行为向量。

$$x_i(k+1)=\begin{cases} norm(x_i(k),\omega e^{(1-\beta(k))}) & x_i(k)\in[x_{min},x_{max}] \\ x_i(k) & 其他 \end{cases} \tag{5}$$

其中：$norm(\mu,\sigma)$ 为正态随机数发生器，其中 μ 为期望，σ 为方差。ω 学习速率因子。图 2 为改进前后算法执行速率对比，由图易知，改进前的 CARLA 在一定次数的学习后，算法执行时间陡然增加，几乎无法运行。改进后的 CARLA 学习次数与算法执行时间为线性关系，与改进前相比，计算时间大大缩短，算法的实时性显著提高。

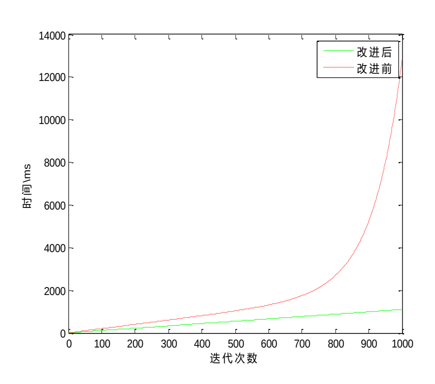


图 2　改进前、后 CARLA 算法执行时间

1.3　PSO 学习环路设计

PSO 最初由 Kennedy 和 Eberhart 提出[16],最初是起源于对简单社会系统的模拟。 PSO 结合了鸟类的认知行为和社会合作。该方法在解决非线性。不可微、多极点、高维问题上具有很好的鲁棒性[17]。PSO 模仿鸟群。鱼群的群集行为，这些群体按照一种合作的方式寻找食物，群体中的成员学习它自身的经验和其他成员的经验不断改变其搜索模式。PSO 同遗传算法类似，是一种基于迭代的优化算法。系统初始化为一组随机解，通过迭代搜寻最优值。PSO 学习环路设计主要有如下几步：

a) 初始化。设置粒子群的大小 N，初始位置 $X_{id}=\{X_{i1},X_{i2}\cdots,X_{in}\}$，初始速度 $V_{id}(0)=\{V_{i1},V_{i2}\cdots,V_{in}\}$，最大速度区间 $[V_{min},V_{max}]$。

b)更新速度和位置公式。

$$V_{id}(k+1)=\delta V_{id}(k)+V_{id1}(k+1)+V_{id2}(k+1) \tag{6}$$

$$X_{id}(k+1)=X_{id}(k)+V_{id}(k+1) \tag{7}$$

$$V^{p}_{id}(k+1)=C_1 random(0,1)(pbest_i(k)-X_{id}(k))$$

$$V^{g}_{id}(k+1)=C_2 random(0,1)(gbest(k)-P_{id}(k))$$

其中：$i=1,2,\cdots,N$，N 为粒子群的总数。$d=1,2,\cdots n$，n 为行为个数，V_{id} 为第 i 个粒子中第 d 个行为的速度，X_{id} 为第 i 个粒子中第 d 个行为的位置。δ 为惯性因子（$\delta>0$），其值越大，全局寻优能力越强，其值越小，局部寻优能力越强。C_1 和 C_2 为称为加速常数，一般的 $C_1=C_2\in[0,4]$，$random(0,1)$ 表示区间 $[0,1]$ 上的均匀随机数，$pbest_i(k)$ 为第 i 个粒子当前最好位置，$gbest(k)$ 为整个粒子群当前的最好位置。

c) 终止条件。有两种终止条件可以选择：一是达到最大代数 G_{max}，二是连续 m 代环境适应度值不在发生变化即停止。

2　应用实例实验

2.1　结晶器液位控制的描述

如图 3 所示，连铸工艺结晶器液位控制流程可简述如下：钢液从钢包（1）通过滑动水口流向中间包（2），控制系统根据浇铸速度通过塞棒（3）的上下移动控制中间包水口的开度。通过控制从中间包流向结晶器（4）钢液的流量来控制结晶器液位的稳定。液位传感器（5）将结晶器液位变化传给控制器（6），控制器驱动塞棒升降机构（7）改变塞棒位置，控制中间包到结晶器的钢液流量，从而达到结晶器液位稳定的目的。

结晶器液位控制具有非线性、时滞、干扰因素多的特点，常规的 PID 控制器很难取得满意的控制效果。在这种情况下吸取智能控制思想并利用计算机技术的优势对传统 PID 控制进行改造，成为当前连铸生产工艺结晶器液位控制研究的热点。本文尝试使用上述组合模型对传统 PID 控制器进行改进，实现结晶器控制器参数在线学习优化。

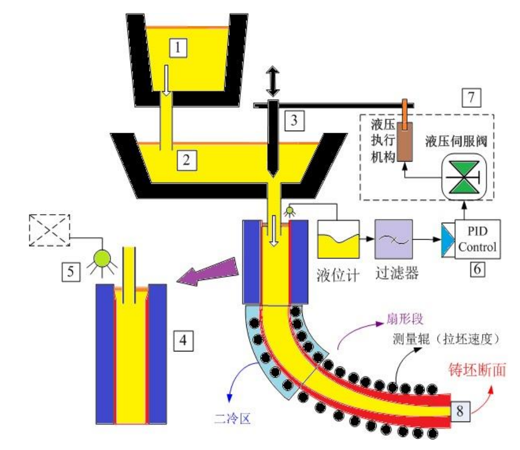


图 3　结晶器液位控制系统结构图

2.2　液位控制模型与参数

本文对唐山某钢厂 2 号连铸机液位控制对象进行系统辨识，主要参数如表 1 所示。为了用参数辨识法获得对象的模型，以液压伺服阀开度作为系统输入变量，结晶器液位为系统输出变量。采用最小二乘法辨识系统模型[18]，所得传递函数为

$$G(s)=\frac{0.08s+1}{0.11s^2+0.1s+0.002}。$$

表 1　液位控制模型参数

参数名称	参数符号	参数取值
拉坯速度	V	$2.3m/\min$
结晶器长度	L	$900mm$
铸坯断面（8）	S	$160mm*160mm$

2.3　环境适应度函数的定义

在设计结晶器液位智能控制器时，首先应该定义控制器性能评价函数（也称环境适应度函数）。典型的时域输出性能指标是系统阶跃响应的超调量、上升时间、稳态时间、稳态误差。总的来讲，常见的控制器性能评价准则有四种：综合绝对误差（IAE，$J=\int|\Delta e|dt$），平方误差积分（ISE，$J=\int(\Delta e)^2dt$），时间加权绝对误差积分（ITAE，$J=\int t|\Delta e|dt$），时间加权平方误差积分（ITSE，$J=\int t(\Delta e)^2dt$）。每一种评价准则都有各自的优缺点。通过最小化 IAE，ISE 能够有效减小超调量，但对缩短稳态时间的影响很小，ITAE、ITSE 可以克服前者的不足，但解析公式的推导比较复杂耗时。考虑到结晶器液位控制实际要求，本文采用 ITSE 评价准则，CARLA-PSO 环境适应度函数定义如下:

$$J^k=\sum_{i=1}^{n}n[Y_i-y_i^{\,k}]^2 \qquad (8)$$

其中：Y 实际响应曲线，y 为理想响应曲线，n 为采样次数，k 为 CARLA-PSO 迭代学习的次数。

2.4　仿真结果分析

依据结晶器液位 PID 控制器实际调参经验，设置 k_p、k_i、k_d 的动作集区间分别为 $[0,2]$，$[0,1]$，$[0,2]$；初始化 CARLA 学习速率因子 ω=0.1；　粒子群大小 N=300，粒子初始速度 $V_{id}(0)=\{0,0\cdots,0\}$，粒子搜索速度区间为 $[-0.5,0.5]$，惯性因子 δ（$\delta>0$）典型取值为 0.4 或 0.9，因为后期为“微调”阶段，所以 δ 取值应较小，取 0.4，C_1= C_2=0.8。结晶器液位控制系统输入信号为单位阶跃信号，结晶器液位采样时间为 50s。

图 4 a) 给出了原系统在无控制器作用下的响应曲线（实线部分），存在明显的超调和稳态误差并且响应时间过长。虚线为理想响应曲线。b) c) d) 给出了具有代表性的三次学习过程，第 4 次迭代优化后适应度值下降了近一半。当迭代优化次数达到 100 次后系统响应曲线无明显超调，响应时间大大缩短，适应度值下降至最初的 10%。此时，优化策略选择器断开 CARLA 学习环路，接通 PSO 学习环路，系统进入后期优化阶段，CARLA-PSO 迭代优化 500 次后系统响应接近理想状态。e) 为 CARLA-PSO 算法收敛曲线。

为了验证 CARLA-PSO 算法的性能，本文给出了 CARLA 与 PSO 算法单独作用下的适应度收敛效果，如图 5 所示。CARLA-PSO 算法结合了 CARLA 全局搜索能力强与 PSO 更新速度快的优点，与 CARLA 算法、PSO 算法相比较，具有更快的收敛速度和更高的优化精度。

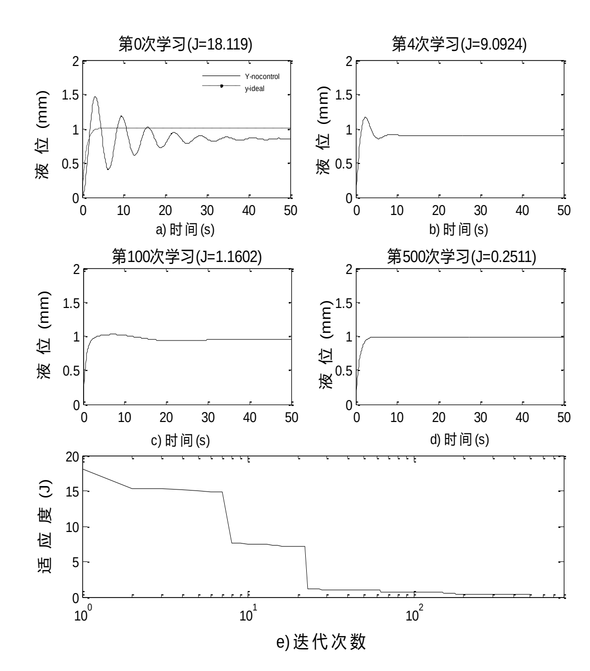


图 4　CARLA-PSO 优化过程示意图

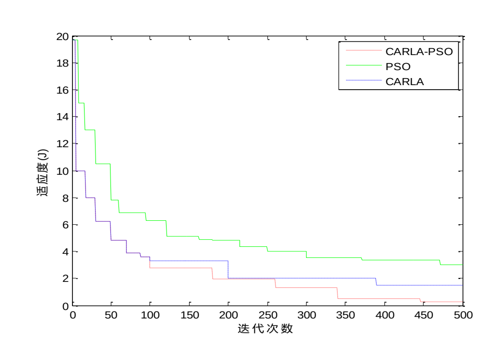


图 5　CARLA-PSO、CARLA、PSO 性能对比

3　结束语

本文讨论了 CARLA 模型，并对传统 CARLA 更新行为函数进行了修改，设计了一种改进的 CARLA，提高了 CARLA 算法的执行速率，增强了算法的实用性。在此基础上，结合 PSO 算法，建立了 CARLA-PSO 组合优化模型。该模型能够自主学习控制规则，收敛速度快，精度高。最后将模型用于连铸结晶器液位控制中，仿真实验表明，在与环境进行多次交互后，模型可自主获取较好的结晶器液位控制规律，与 CARLA 单独作用于系统相比，CARLA-PSO 收敛速度快，优化效率高。对于该模型下一步研究的重点是连铸其他控制环节，并用于真实的结晶器液位控制系统或其他被控对象。

参考文献：

[1]　李祖林. 结晶器液位专家系统控制研究 [J]. 计算机工程与应用, 2009, 45 (10): 199-201.

[2]　Passenbrunner T. Mold level control of a continuous casting plant by

switching control strategies [J]. IFAC Proceedings Volumes, 2010, 43 (14): 1350-1355.
[3] 童朝南，肖磊，彭开香，等. 基于遗传算法的结晶器液位约束广义预测控制 [J]. 控制与决策, 2009, 24 (11): 1735-1739.
[4] 曹光明，吴迪，张殿华. 基于模糊自适应 PID 的铸轧机结晶器液位控制系统 [J]. 控制与决策, 2007, 22 (4): 399-402.
[5] 房启超，徐林，王建辉，等. 改进的 PSO 及其在结晶器液位控制中的应用 [J]. 仪器仪表学报, 2006, 27 (11): 1399-1402.
[6] Delgarm N, Sajadi B, Kowsary F, et al. Multi-objective optimization of the building energy performance: a simulation-based approach by means of particle swarm optimization (PSO) [J]. Applied Energy, 2016, 170: 293-303.
[7] 熊伟丽，徐保国，周其明. 基于改进粒子群算法的 PID 参数优化方法研究 [J]. 计算机工程, 2005, 31 (24): 41-43.
[8] 黄为勇，徐晓菊，潘晓博，等. 量子粒子群优化算法的收缩—扩张系数控制策略研究 [J]. 计算机应用研究, 2016, 33 (9): 2592-2595.
[9] 刘晓. 基于连续动作学习自动机的联想强化学习 [J]. 山西大学学报: 自然科学版, 2015, 38 (3): 426-431.
[10] 周锐，陈宗基. 强化学习在导弹制导中的应用 [J]. 控制理论与应用, 2001, 18 (5): 748-750.
[11] 钟宇平，王丽丹，段书凯，等. 基于神经网络及强化学习的智能控制系统 [J]. 西南大学学报: 自然科学版, 2013, 35 (11): 172-179.
[12] Shah H, Gopal M. A reinforcement learning algorithm with evolving fuzzy neural networks [J]. IFAC Proceedings, 2014, 47 (1): 1161-1165.
[13] Mirsaleh M R, Meybodi M R. A learning automata-based memetic algorithm [J]. Genetic Programming & Evolvable Machines, 2015, 16 (4): 399-453.
[14] Howell M N, Gordon T J, Best M C. The application of continuous action reinforcement learning automata to adaptive PID tuning [C]// Learning Systems for Control. London: IEEE Xplore, 2000: 2//1-2//4.
[15] Kashki M, Abdel-Magid Y L, Abido M A. A reinforcement learning automata optimization approach for optimum tuning of PID controller in AVR system [C]// Proc of International Conference on Intelligent Computing: Advanced Intelligent Computing Theories and Applications-with Aspects of Artificial Intelligence. [S. l. [: Springer-Verlag, 2008: 684-692.
[16] Hasanien H M. Design Optimization of PID controller in automatic voltage regulator system using taguchi combined genetic algorithm method [J]. IEEE Systems Journal, 2013, 7 (4): 825-831.
[17] 宋明智，杨乐. 基于改进自适应 PSO 算法的 WSN 覆盖优化方法 [J]. 计算机应用研究, 2013, 30 (11): 3472-3475.
[18] 章家岩，郎佳红，王金忠. 转炉炉口微差压滑模变结构控制技术 [J]. 控制工程, 2009, 16 (4): 419-422.